# Investigating the asymmetry of English sibilant assimilation: Acoustic and EPG data

MARIANNE POUPLIER*, PHILIP HOOLE*, and JAMES M. SCOBBIE**

**Institute of Phonetics and Speech Processing, Ludwig-Maximilians University, Germany*
***CASL, Queen Margaret University, Scotland*

*Abstract*

*We present tongue-palate contact (EPG) and acoustic data on English sibilant assimilation, with a particular focus on the asymmetry arising from the order of the sibilants. It is generally known that /s#ʃ/ sequences may display varying degrees of regressive assimilation in fluent speech, yet for /ʃ#s/ it is widely assumed that no assimilation takes place, although the empirical content of this assumption has rarely been investigated nor a clear theoretical explanation proposed. We systematically compare the two sibilant orders in word-boundary clusters. Our data show that /s#ʃ/ sequences assimilate frequently and this assimilation is strictly regressive. The assimilated sequence may be indistinguishable from a homorganic control sequence by our measures, or it can be characterized by measurement values intermediate to those typical for /ʃ/ or /s/. /ʃ#s/ sequences may also show regressive assimilation, albeit less frequently and to a lesser degree. Assimilated /ʃ#s/ sequences are always distinguishable from /s#s/ sequences. In a few cases, we identify progressive assimilation for /ʃ#s/. We discuss how to account for the differences in degree of assimilation, and we propose that the order asymmetry may arise from the different articulatory control structures employed for the two sibilants in conjunction with phonotactic probability effects.*

## 1. Introduction

One of the most extensively discussed phenomena of connected speech is the assimilation of a word-final consonant to a following word-initial consonant, because this tendency of final consonants to lose their identity can be used to explore topics such as the cross-linguistic distribution of phonological contrast, phonetic planning, and prosodic structure. However, not all final consonants act alike. There are systematic differences in assimilation patterns depending on factors such as the place or manner of the word-final consonant, and indeed the word-initial one. Even consonants as similar as /s/ and /ʃ/ can be expected to behave differently in word-final position. Quite how and why this is so remains contentious, and our current

JLP 2 (2011), 1–33
DOI 10.1515/LABPHON.2011.001
1836-6346/11/01–1

study adds to the assimilation literature by investigating in detail how English sibilant assimilation differs as a function of sibilant order. It is a well known phenomenon that /s#ʃ/ sequences in English, such as in the phrase *Paris show*, frequently assimilate in fluent speech to a form similar to [ʃʃ], so this ordering of the sibilants has been well studied. But because it is believed that the opposite order of sibilants, /ʃ#s/ (e.g., *fish soup*), does not assimilate in the same fashion, or at all, this ordering has only rarely been discussed. Indeed, it is simply not known whether there is any significant assimilation at all in the latter case, and if so, how it patterns.

A further goal of our study is to shed light on the cognitive and physical aspects of assimilation. English sibilant assimilation has featured prominently in the debate on how cognitive planning of spoken language should be characterized to account for the variety of assimilation patterns observed. Two overall different approaches to assimilation are thereby usually pitted against each other. In nonlinear approaches to phonology such as Autosegmental Phonology or Feature Geometry, assimilation occurs when a distinctive feature (or subset of features) within a segment changes to agree with the feature(s) of an adjacent segment. This is achieved through linking and de-linking of features (Goldsmith 1976; Clements 1985; McCarthy 1988). In this type of model, assimilation is by definition a categorical replacement process which happens prior to and independently of the computation of the physical properties of the utterance. Articulatory and acoustic recordings of assimilated sequences have shown, however, that this is not necessarily correct: For example, in English [ʃ#j] sequences such as *miss you* can be pronounced as [mɪʃju] in connected speech. Articulatorily and acoustically, the assimilated [ʃ] differs from an underlying /ʃʃ/. This has been taken to mean that this [ʃ]-like pronunciation is caused by the tongue tip gesture for /s/ and the tongue body raising gesture for /j/ sliding into each other; the gestures temporally overlap. The assimilated fricative is then a blend of the two overlapping targets /s, j/ (Browman and Goldstein 1990b; Zsiga 1995). For lexically derived forms, however, such as *impression*, Zsiga finds no evidence of the sibilant arising from a coproduction of /s#j/ gestures. She therefore proposes that assimilation in lexically derived forms arises through symbolic (de)linking of features, while postlexical assimilation arises from gestural overlap (for further discussion see also Scobbie 1995).

On the gestural overlap account of assimilation as first proposed by Browman and Goldstein within their Articulatory Phonology framework (Browman and Goldstein 1990a, 1992; Goldstein, Byrd, and Saltzman 2006), fluent speech phenomena such as (apparent) deletions, assimilation, and weakening are all traced back to a single underlying principle: different degrees of gestural overlap, which may or may not be accompanied by a partially reduced articulation of the overlapped gesture. Depending on the particular articulators employed by the overlapping gestures, different consequences are observed. The gestural approach therefore predicts precisely the intermediate and gradient renditions of assimilated sequences that the autosegmental approach is not able to account for in a

straightforward fashion. However, other data pose greater problems for the gestural model. Several studies have shown that assimilation can be complete[1]: The assimilated gestures may be consistently not produced, as predicted by the symbolic (de-)linking account. For example, for Korean, Son et al. (2007) showed that in word-medial /pk/ clusters, the /p/ is either fully present or not produced at all (but see Jun 1996). Similarly, the tongue tip gesture for word-final alveolar /n/ in Castilian Spanish (e.g., *diga[n] → diga[m] paja*) is fully reduced and the lip aperture gesture is temporally extended (Honorof 1999). Yet matters are more complex than that, since complete assimilation only occurs when there is a following noncoronal; for following coronals, a blended output is observed, in line with the gestural overlap approach. Also Korean coronals in coronal+stop clusters (/tp, tk/) were shown to reduce completely in most, but gradiently in some of the cases, with the occurrence of reduction varying as a function of speaking rate (Kochetov and Pouplier 2008). While cases of consistent and complete assimilation might be understood from a diachronic perspective in that a formerly gradient, fluent-speech assimilation process has become lexicalized and is independent of postlexical factors such as speech rate which usually condition gradient assimilation, there are studies demonstrating that such an account falls short of a complete explanation. Complete and gradient assimilations coexist in connected speech for the same lexical items and for the same speaker. Ellis and Hardcastle (2002) investigated /n#k/ sequences in English and found that some speakers produced an assimilatory continuum between [nk] and [ŋk], while others showed a binary opposition between unassimilated [nk] or fully assimilated [ŋk], with no evidence for a nonvelar target contributing to the output articulation. Other studies describe a similar range of speaker behaviour, and assimilation patterns consonant with a symbolic linking-delinking view have emerged side-by-side with gradient assimilation and gestural hiding and blending phenomena (for example, Barry 1991; Nolan 1999; Kühnert and Hoole 2004; Kochetov and Pouplier 2008).

In this context, the particular characteristics of /s#ʃ/ assimilation have received considerable attention in an exchange between Nolan et al. and Browman (Nolan 1992; Browman 1995; Nolan, Holst, and Kühnert 1996), see also Zsiga (1995). Nolan et al. have contested Browman and Goldstein's claim that assimilation in connected speech arises exclusively from the spatio-temporal overlap of gestures. Nolan et al. demonstrated that English /s#ʃ/ assimilation can in some (but not all) cases render a categorical assimilation to /ʃ#ʃ/. They assume this assimilation to be categorical because an analysis of the assimilated fricative sequence in terms of tongue-palate contact data (EPG) does not show any trace of a blended /s#ʃ/ articulation, as would be expected from overlapping articulatory gestures and is indistinguishable from an underlying /ʃ/. There may be other differences that were not measured, but since the theoretical debate is at the level of the coronal place feature or gesture, this point is generally not explored further. Indeed, the authors place a lot of emphasis on the fact that the global acoustic duration of these assimilated sibilant sequences is significantly longer than that of a corresponding

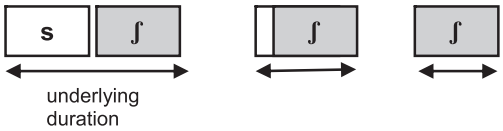


Figure 1. *Schematic illustration of how assimilation arising from gestural overlap may lead to a shortened duration of the assimilated sequence in the gestural score.*

singleton sibilant in a nonassimilating control sequence. Therefore, so the authors conclude, the assimilation in these cases has to arise from a cognitive restructuring of discrete symbolic representations, not from the physical overlap of articulations. This interpretation is somewhat problematic because, among other reasons, it cannot be excluded that the physical overlap may only be complete in terms of tongue-palate contact, and incomplete elsewhere in ways not measurable with EPG. Further, it is not obvious to what extent durational differences between the assimilated sequence and a control singleton should be used to come to conclusions about the nature of the assimilation mechanism. From a gestural perspective, the general premise is indeed that, all else being equal, the duration of a cluster should decrease with increasing assimilation (i.e., increasing overlap). Why this is so is schematically illustrated in Figure 1.

Yet the problem with surface duration measurements is, of course, that 'all else being equal' may very well not hold, and any number of factors may contribute to differences in duration between non-assimilating /ps#p/ or /p#ʃ/ (*"claps Paul", "clap Shaun"*) and assimilated /ps#ʃ/ (*"claps Shaun"*) sequences. Whatever the details of the phonological formalism, it is usually assumed that a limited number of categorically-distinct options are available from a symbolic, linking and delinking of features perspective (though see Scobbie 1995). Under the simplest account, the spreading feature comes to occupy two symbolic timing slots, hence a longer duration of an assimilated sequence would be predicted.[2] However, the number of factors influencing the surface duration of a segment or sequence of segments are many. Subsequent phonetic studies have borne out that feature-spreading is an over-simplistic analysis if the full range of assimilation patterns is to be accounted for, and have found support for a more complex range of gestural behaviours. Jun (2004), for instance, postulates in his account of Korean /tk → kk/ place assimilation that gestural reduction is accompanied by temporal extension of the unreduced velar gesture. Kochetov and Pouplier (2008) confirm that in Korean assimilated /kk/ sequences (with no residual tongue tip gesture), the dorsal gesture may be extended in time and have a similar duration to an underlying /kk/ sequence. However, for assimilated /tp → pp/ sequences, the assimilated /pp/ sequence was shorter compared to an underlying /pp/ sequence. Kühnert and Hoole (2004) find that underlying sequences of two identical segments may be as short as singleton consonants; hence it is unclear whether the duration of a singleton consonant is an appropriate reference duration from which the underly-

ing assimilation mechanisms can be inferred. Nolan et al. (1996) focused their theoretical conclusions exclusively on the existence and characteristics of completely assimilated tokens (since those were their main point of interest at the time), and only reported results from those tokens. However, a highly interesting aspect neglected in that debate is the coexistence of gradient *and* categorical effects in assimilation, since neither a feature spreading nor a gestural overlap account straightforwardly accommodates the range of phenomena observed. Completely assimilated tokens form part of a distribution, and one question to be answered is whether such distributions arise from two discrete and distinct mechanisms (phonetic and phonological) or from the varied and potentially discontinuous parameterisation of a single mechanism, which may perhaps be seen as spanning both theoretical levels.

Another interesting aspect of this class of assimilations is the manner asymmetry: Coronal stops are generally more likely to assimilate, while fricatives are less likely to assimilate. Neither a gestural overlap nor a nonlinear phonological account of place assimilation makes inherent predictions about why manner matters. Much attention has been paid in the literature to differences between consonant classes, especially on the 'special' status of coronal stops as targets, but not triggers of place assimilation. Word-final coronals are likely to assimilate in place to a following non-coronal stop, but word-final non-coronals do not assimilate to a following coronal stop. Also in this context, sibilants present a very interesting case: In English, /s, ʃ/ are both considered to be coronal fricatives, phonologically at least. They both are very similar in place and manner and yet show differences in coarticulatory behaviour. This is quite different from the case of stops, for which the different places of articulation that show assimilatory asymmetries are spaced quite far apart in the vocal tract.[3]

Asymmetries in assimilation patterns have been approached from two overall perspectives, one articulatory (resistance to coarticulation), the other perceptual, both of which we shall briefly summarize in turn. In the resistance to coarticulation research paradigm (Recasens, Fontdevila, and Pallarès 1992; Recasens 2006; Recasens and Espinosa 2009), differences in coarticulatory and assimilatory behaviour have been attributed to differences in tongue body control. This latter factor is seen as the main predictor of the extent to which consonants are influenced by their neighbours. Those consonants that do not control the tongue body, such as coronals, are least resistant to coarticulation and exert the least coarticulatory influence on neighbouring consonants (they show the least coarticulatory aggression). Also manner and any conflicting demands on the articulators involved may impact coarticulation resistance. Work by Recasens et al. (1997) shows that both sibilants, although differing in tongue body articulation, are highly coarticulation resistant, presumably due to their manner. Given these results, an asymmetry between /s#ʃ/ and /ʃ#s/ is unexpected.

Another account for asymmetries in assimilation patterns has been put forward from a perceptual perspective (among others, Kohler 1990; Ohala 1990; Jun 1995;

Steriade 2001, 2009). It has been proposed that assimilation is most likely for the least perceptible consonants in perceptually weakest structural positions. Fricatives do not assimilate where stops do because fricatives are perceptually more salient. Also the regressive nature of assimilation is seen as falling out from perceptual factors, since word-final sounds are less perceptually salient compared to word-initial sounds. Their importance for lexical access may further stabilize word-initial sounds. None of the perceptual accounts of assimilation tackle the asymmetries between /s/ and /ʃ/ though, since both consonants are assumed to be perceptually among the most salient segments.

We are aware of only two published studies that explicitly investigate both orders of sibilants. Perkell et al. (Perkell, Boyce, and Stevens 1979) report, in a short communication, a study of both /ʃ#s/ and /s#ʃ/ sequences. The authors employ acoustic measurements as well as a measurement of tongue tip contact with a pair of electrodes fixed to the inner aspect of the lower alveolar ridge. Contact can be observed in /s#s/ but generally not for /ʃ#ʃ/ control sequences (cf. also Perkell et al. 2004). In terms of these measures, /ʃ#s/ sequences can clearly be identified as a sequence of two sibilants, but /s#ʃ/ sequences can be identical to /ʃ#ʃ/ sequences. The only other published study that investigates both /s#ʃ/ and /ʃ#s/ sequences is by Niebuhr et al. for French (Niebuhr, Lancia, and Meunier 2008). They report that, similarly to English, there is a bias towards /ʃ/ in sibilant assimilation, but French shows a /ʃ/ bias in both /s#ʃ/ and /ʃ#s/ sequences. That is, assimilation takes either progressive or regressive directionality so that it always leads to assimilation towards /ʃ/, although regressive assimilation is stronger than progressive assimilation.

In sum, a detailed study of sibilant assimilation including both /s#ʃ/ and /ʃ#s/ sequences is an ideal testing ground for different theoretical accounts of assimilation. The goals of our current study are thus twofold. For one, we focus on the apparent asymmetry of assimilation conditioned by the order of the sibilants. We compare the assimilatory patterns for /s#ʃ/ and /ʃ#s/ sequences and investigate whether /ʃ#s/ sequences show any signs of regressive (or, potentially, progressive) assimilation. Secondly, our data speak to the debate of gradient versus complete nature of assimilation by offering a more comprehensive analysis of English sibilant assimilation compared to earlier studies.

## 2. Method

### 2.1. *Experimental setup and design*

We recorded electropalatographic (EPG) tongue-palate contact data sampled at 200 Hz by means of an Articulate Instruments multichannel WinEPG system. Acoustic data were simultaneously recorded and sampled at 48 kHz. We recorded ten speakers, all of them staff at Queen Margaret University, Edinburgh, and in possession of an individually-fitted EPG palate with 62 electrodes. All are native

Table 1. *Stimuli recorded for the present experiment.*

| Condition | Cluster | Set | | |
|---|---|---|---|---|
| | | PARIS/PARISH | MISS/FISH | CAPS/CASH[4] |
| heterorganic conditions | *s#ʃ* | Paris show | Miss Shoe | caps shares |
| | *ʃ#s* | parish soap | fish soup | cash Sarah's |
| homorganic conditions | *s#s* | Paris soap | Miss Soup | caps saplings |
| | *ʃ#ʃ* | parish show | fish shoe | cash shares |

speakers of English as spoken in Great Britain. The speakers differed as to whether they spoke Standard Southern British English (SSBE, "EN" subjects) or Scottish-accented Standard English (SSE, "SC" subjects), even though dialectal differences were not explicitly targeted by our experimental design. All of our speakers were very experienced in speaking with an EPG palate; they also wore their EPG palates for at least half an hour before the recording time. One speaker's palate had a different design (SC5) from the others, but since we are only comparing relative differences to each individual speaker's typical /s/ or /ʃ/ patterns, this does not affect our analyses.

Speakers were familiarised with the stimulus sentences before the recording session. Sentences were presented on a screen, one at a time, in four pseudo-randomised blocks, the same for each subject. Our speakers were naïve as to the research questions addressed by the experiment, with the exception of SC1, the last author. Materials consisted of words ending and beginning in the sibilants /s/ and /ʃ/, rendering two homorganic (/s#s/, /ʃ#ʃ/) and two heterorganic cluster conditions (/s#ʃ/, /ʃ#s/). The sibilant sequences were recorded in three different "sets" (see Table 1) which differed in the quality of the surrounding vowels. The target words were embedded in carrier phrases, such as "The Paris show is now open." The sentences were embedded in stimulus material for a different experiment focusing on laterals that was recorded at the same time (Scobbie and Pouplier 2010).

Overall, the experiment was designed to elicit 48 tokens per subject, with four sequences (/s#s/, /ʃ#ʃ/, /s#ʃ/, /ʃ#s/) × three sets (*CAPS*|*CASH*, *MISS*|*FISH*, *PARIS*|*PARISH*) × four repetitions. Due to technical problems during data collection there are slightly fewer tokens for some speakers; notably SC1 completed only three repetitions. Overall 17 tokens were lost, leaving a total of 463 tokens across subjects.

## 2.2. *Data analysis methods*

The acoustic signal was hand segmented for the entire intervocalic sibilant interval for all conditions. All EPG and acoustic analyses are based on these acoustic sibilant intervals. We automatically calculated the 25%, 50%, and 75% time points for

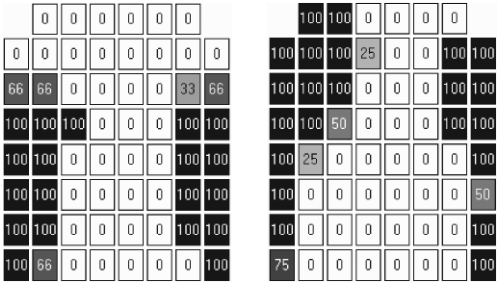


Figure 2. *Average EPG contact pattern at the acoustic midpoint for speaker EN1, for the homorganic conditions from the PARIS|PARISH set. |ʃ#ʃ| is on the left, |s#s| on the right. Numbers present percent contact for each electrode.*

each interval and used these as primary analysis timepoints, but we also present dynamic analyses across the sibilant intervals.

To analyse the EPG data we developed a normalization procedure based on Gusik and Harrington (2007). We created a normalized EPG index so that each palatogram could be evaluated on a scale ranging from −1 to 1, with −1 being the most extreme value for a typical /ʃ/ and 1 being the most extreme value for a typical /s/. This procedure was carried out as follows, separately for each subject and each set (to take into consideration the different vowel environments). For each speaker we calculated the (across-repetition) average contact per electrode at the acoustic midpoint of the control conditions /ʃ#ʃ/ and /s#s/. This rendered a typical /s/ and a typical /ʃ/ contact pattern for a given speaker and set, based on the four repetitions of each control sentence; see Figure 2 for an example. A difference pattern was then created by subtracting for each electrode the value of the /ʃ/ pattern from the value of the /s/ pattern. For each electrode that displays a high contact value for /s/ but not for /ʃ/, a large positive number will result, and a large negative number in cases in which the contact for /ʃ/ substantially exceeds the one for /s/. For contacts that are very similar or the same in /s/ and /ʃ/, a value of zero or a small value (positive or negative) will result in the corresponding cell in the difference pattern. The entire difference pattern was then normalized by dividing each cell's value in the pattern by the sum of the absolute values of the pattern as a whole. For each target palatogram (e.g., the palatogram at the 25% timepoint of the /s#ʃ/ condition of a particular token), a sum was computed over all electrodes in which each contacted electrode contributes its value in the difference pattern directly, and in which each uncontacted electrode contributes the negative of its value in the difference pattern. The result is, for each EPG sample, a value between −1 and 1 with −1 being maximally ʃ-like and 1 being maximally s-like.

In order to quantify differences in assimilatory patterns, we use the distributional properties of the homorganic conditions to identify 'gradient' and 'complete' assimilations at the 25% timepoint. Any token from the heterorganic C1#C2

condition more than two standard deviations from the corresponding homorganic C1 control mean is identified as assimilated. If tokens are more than 2 SDs from both (C1 and C2) control means they are classified as gradiently assimilated. If they are more than 2 SDs from their appropriate C1 control mean, and less than 2 SDs from the other (C2) control mean, they are considered completely assimilated (see endnote 2 on the meaning of complete). The means and their SDs were computed separately for each speaker, collapsing across sets. Therefore the mean and SD are computed based on 12 tokens for each of the sibilants per subject. The 2SD threshold is of course an operational criterion only. It serves as a classification tool that enables us to quantify differences in articulatory patterns across subjects and conditions; see Kochetov and Pouplier (2008) for a similar procedure.

As is well known, sibilants show differential characteristics in the spectral properties of the frication noise. /ʃ/ is characterized by a concentration of energies at lower frequencies in the spectrum due to the more posterior constriction, possibly enhanced by lip rounding. For the acoustic analysis we calculated spectra for each cluster (continuously and at each analysis timepoint) using the multi-taper method (Thomson 1982; see Chitoran and Iskarous 2008 for an application to speech data). It is a known problem for the analysis of fricatives that there is a high variance in the spectral shape from one window to the next due to the stochastic quality of the fricative noise. We decided to alleviate the problem using a Matlab implementation of Thomson's multi-taper method. Instead of calculating a spectrum from a single time window, this method uses the weighted average of, in our case, seven orthogonal windows (discrete prolate spheroidal sequences). In the Matlab function *pmtm* this corresponds to the default value of 4 for the time-bandwidth product. The audio data were downsampled to 24 kHz. The window length used was 21.3 ms (512 points at 24 kHz) and was moved across the entire utterance with 75% overlap. This method is equivalent to using a 375 Hz filter bandwidth.

The quantification of the spectral differences between the sibilants proceeded in analogy to the calculation of the normalised EPG index (each EPG frame can be considered as a vector of 62 values, just as each spectrum consists of a vector of, e.g., 256 values), but differed in the arithmetic details because for spectra it is not possible to exploit the restriction of EPG patterns to values of 0 and 1. First, reference patterns for /s/ and /ʃ/ were computed on the basis of spectra taken at the temporal midpoint of the homorganic control conditions. For the following index calculation the frequency region from 1 kHz to 11 kHz was used. For each timepoint and utterance the Euclidean distance of the spectral slice from each of these two reference patterns was calculated. To reduce the influence of overall changes in the signal level, the average value over the given frequency region was subtracted from the measured and reference spectra before computing these Euclidean distances. The final index value was then computed by dividing the Euclidean distance from the ʃ-reference by the sum of the two Euclidean distances, subtracting 0.5 and multiplying by 2, giving as for the EPG index values ranging from −1 (most /ʃ/ like) to +1 (most /s/ like).[5]

Table 2. *Means and SDs for the EPG index of the homorganic conditions.*

| Set | Timepoint | ʃ#ʃ | | s#s | |
|---|---|---|---|---|---|
| | | Mean | SD | Mean | SD |
| CAPS/CASH | 25 | −.79 | .03 | .88 | .02 |
| | 50 | −.86 | .02 | .90 | .01 |
| | 75 | −.87 | .03 | .82 | .05 |
| MISS/FISH | 25 | −.77 | .03 | .78 | .03 |
| | 50 | −.88 | .01 | .88 | .01 |
| | 75 | −.85 | .02 | .84 | .03 |
| PARIS/PARISH | 25 | −.79 | .03 | .69 | .04 |
| | 50 | −.87 | .01 | .87 | .02 |
| | 75 | −.84 | .02 | .89 | .02 |

## 3. Results

### 3.1. *Spatial index analysis*

Our primary question concerns how the two consonants influence each other depending on their order. In order to adequately assess whether the heterorganic conditions show assimilation, it is important to consider the range of variability exhibited by the homorganic cluster conditions with respect to our measures employed. For the EPG index measure, Table 2 gives the means and standard deviations for the homorganic conditions by set at the three analysis timepoints. As a reminder, we expect values approaching 1 for /s/ and values approximating −1 for /ʃ/. The sibilants are clearly separated by our analysis method, and the values are as expected. A repeated measures ANOVA with the factors Cluster (/s#s/, /ʃ#ʃ/), Set (CAPS/CASH, MISS/FISH, PARIS/PARISH) and Timepoint (25, 50, 75) showed significant differences for the main factor Cluster ($F(1,9) = 3546.3$, $p < .01$), but not for the factors Set ($F(2,18) = 1.78$, $p = .2$) and Timepoint ($F(2,18) < 1$). The ANOVA confirms that our EPG index difference metric is successful in separating out the two sibilants and that the data can be collapsed across sets. Where informative, by-set data will still be given. We also tested for dialectal differences in the control conditions, collapsed across analysis timepoints and sets, for our two dialectal speaker groups. Mann-Whitney-U tests conducted separately for the two sibilants were not significant (/s/: $Z = -.313$, $p = .75$; /ʃ/: $Z = -.522$, $p = .602$); therefore, dialect will not be a factor in our further analyses.

Figure 3 gives the histograms for the EPG index data at all three timepoints, collapsed across sets and across subjects. The data at each analysis timepoint were grouped into 10 equally spaced histogram bins. The values on the x-axis indicate the centres of the bins. Note that the controls are identical in the left and right columns, since for both sibilant sequences, the same /s#s/ and /ʃ#ʃ/ sentences were used as controls.

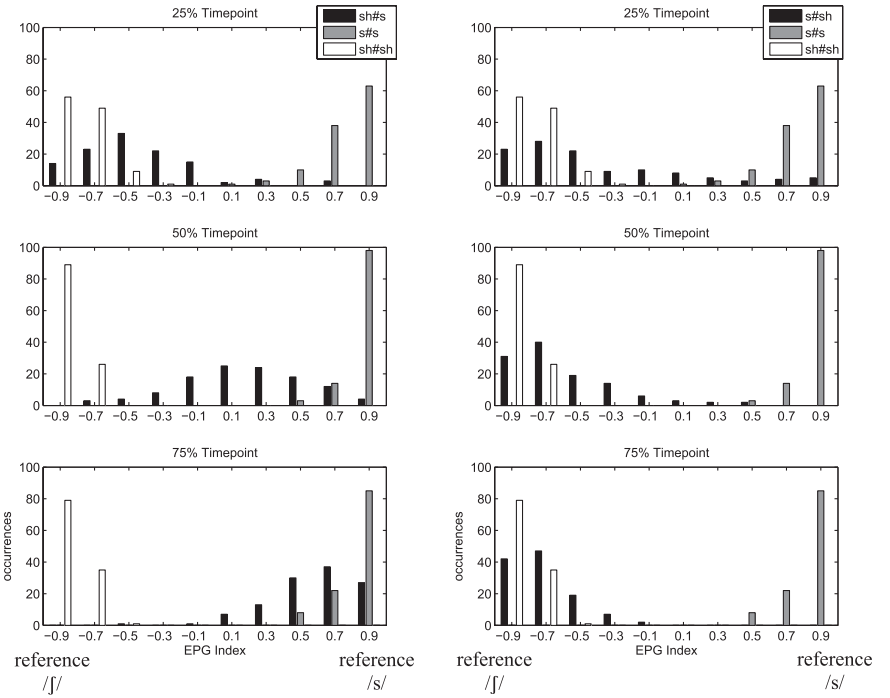


Figure 3. *Comparison of the EPG index for the heterorganic clusters and the homorganic control conditions. /ʃ#s/ is compared in the left panels, /s#ʃ/ in the right. The three analysis timepoints are shown top to bottom. EPG index values are allocated to 10 equally spaced bins. Tick labels on the x-axis indicate bin centres.*

We first concentrate our analysis on the 25% timepoint. At the 25% timepoint for a /ʃ#s/ sequence (top left panel of Figure 3), we expect, if there is no assimilation, values approximating −1, since the /ʃ/-portion of the heterorganic fricative sequence can be expected to be /ʃ/-like. Vice versa, for a /s#ʃ/ sequence (top right panel of Figure 3), we would for unassimilated sequences expect values approximating 1 at the 25% timepoint, since this corresponds to the /s/-portion of the fricative sequence. Table 3 shows the means for the four conditions for the 25% timepoint across subjects.

At the 25% timepoint, we see that /ʃ#s/ tokens are closer to the /ʃ#ʃ/ tokens, although the histogram distribution (Figure 3) extends more to the right. On average, /ʃ#s/ indices are negative, even though clearly distinct from the homorganic /ʃ#ʃ/ cluster. Few tokens make contact with the edges of the /s#s/ distribution. The distribution of the /s#ʃ/ tokens looks remarkably similar showing high degrees of early regressive assimilation. The majority of tokens are closer to the /ʃ#ʃ/ distribution rather than the positive values characteristic for the /s#s/ distribution. Also the mean EPG index is very close in value to the one seen for /ʃ#s/. A repeated

Table 3. *Means and SDs for the EPG index at the 25% timepoint by cluster.*

| | EPG Index at the 25% Timepoint | |
|---|---|---|
| Cluster | Mean | SD |
| ʃ#s | −.47 | .33 |
| ʃ#ʃ | −.78 | .14 |
| s#s | .80 | .15 |
| s#ʃ | −.38 | .49 |

measures ANOVA was conducted with the factor Cluster containing the four levels /s#s/, /ʃ#ʃ/, /s#ʃ/, and /ʃ#s/. The main effect was significant ($F(3,27) = 116.5$, $p < .01$). Bonferroni-corrected comparisons of the means show that /s#ʃ/ and /ʃ#s/ do not differ significantly from each other at the 25% timepoint. /s#ʃ/ differs significantly from its control condition /s#s/ ($p < .001$), but also differs from /ʃ#ʃ/ ($p = .021$). Likewise /ʃ#s/: The cluster differs significantly from both /s#s/ ($p < .001$) and /ʃ#ʃ/ ($p = .001$). Overall, at 25% /ʃ#s/ is still relatively closer to the /ʃ#ʃ/ distribution, but it nonetheless differs from the control condition early on in the sibilant cluster, displaying assimilation. /s#ʃ/ displays strong regressive assimilation towards /ʃ/ at the 25% timepoint. /ʃ#s/ displays some regressive assimilation, albeit to a lesser degree.

Categorizing tokens operationally based on our assimilation metric, using index thresholds of 2 standard deviations (see Methods), we obtain the results displayed in Figure 4. Only assimilated tokens are included in the figure. The results are given in percentage of tokens per subject falling into the gradient and complete categories (recall that some subjects recorded fewer repetitions than others). If all tokens were recorded successfully, 100% represents 12 tokens, as detailed in the Methods section. We will present other, non-categorical analyses of assimilation degree and the dynamics of change further below.

All speakers show at least some assimilated tokens, and in both sibilant orders. /ʃ#s/ sequences differ from /s#ʃ/ sequences, because the former show almost exclusively gradiently assimilated tokens, and overall a lower rate of assimilation. Just three /ʃ#s/ tokens (3%) are completely assimilated (EN3, SC2), while 58% of tokens are gradiently assimilated. /s#ʃ/ shows on average 48% gradient assimilation, and a remarkable 44% of tokens fall into the complete assimilation category. There are some interspeaker differences, however: EN3 shows almost exclusively complete assimilation while EN1 and EN2 show gradient assimilation only. A chi-square test was conducted to test whether the type of assimilation observed depends on the order of sibilants. The contingency table for the percent data across subjects is given in Table 4. The chi-square test was significant ($\chi^2(2) = 29.4$, $p < .001$). The percentages in Table 4 indicate that the significant result stems from the differences in the complete and no assimilation categories, in which the two sibilant orders exhibit almost opposite behaviour.

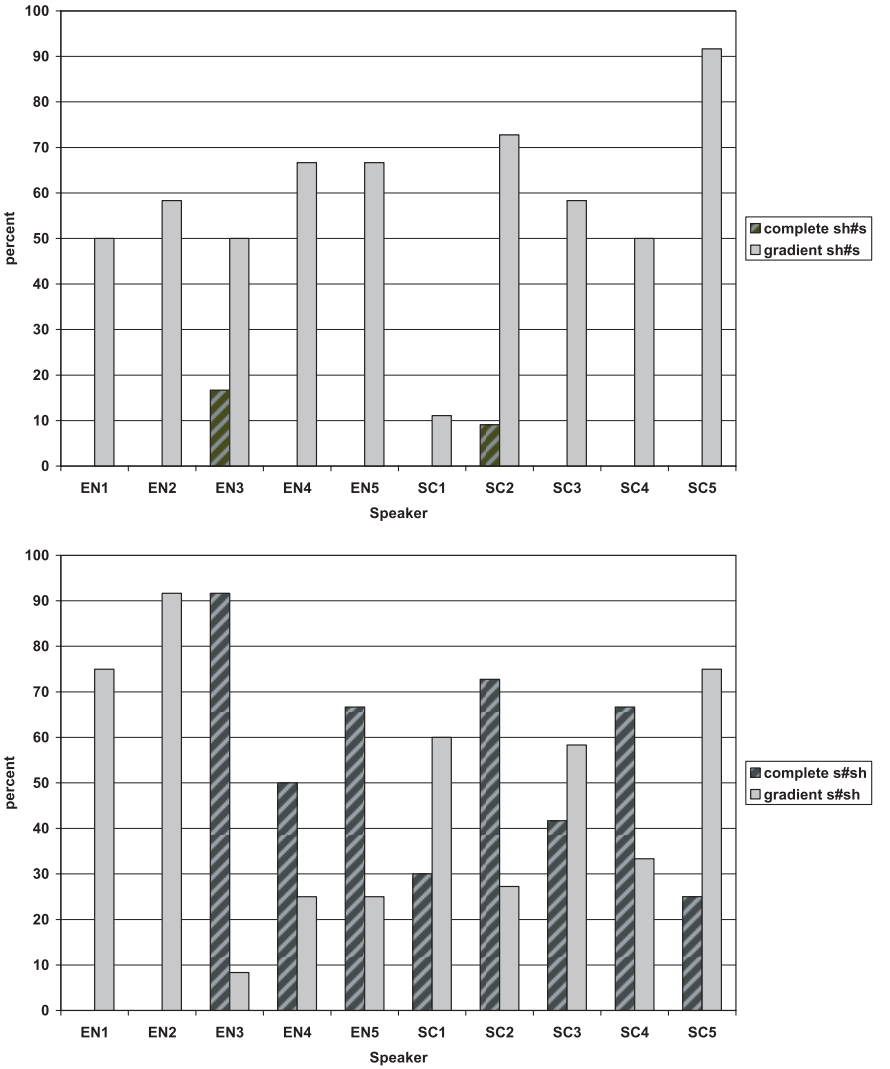


Figure 4. *Percentage of tokens categorized as complete (striped bars) or gradient (filled bars) assimilations. |ʃ#s| sequences are at the top, |s#ʃ| sequences at the bottom. Unassimilated tokens are not shown; therefore, percentages do not add up to 100.*

Table 4. *Contingency table for assimilation type and sibilant order.*

| order | % gradient | % complete | % no assimilation |
|---|---|---|---|
| s#ʃ | 24 | 22 | 4 |
| ʃ#s | 29 | 1 | 19 |

Table 5. *Normalized spectral difference values (Acoustic index) for the homorganic clusters by set.*

| Set | Timepoint | ʃ#ʃ | | s#s | |
|---|---|---|---|---|---|
| | | Mean | SD | Mean | SD |
| CAPS/CASH | 25 | −.49 | .12 | .40 | .15 |
| | 50 | −.57 | .10 | .54 | .08 |
| | 75 | −.50 | .09 | .43 | .13 |
| | Average | −.52 | .11 | .46 | .14 |
| MISS/FISH | 25 | −.45 | .12 | .43 | .15 |
| | 50 | −.57 | .12 | .53 | .13 |
| | 75 | −.49 | .12 | .45 | .12 |
| | Average | −.51 | .13 | .47 | .14 |
| PARIS/PARISH | 25 | −.48 | .13 | .43 | .14 |
| | 50 | −.59 | .13 | .53 | .08 |
| | 75 | −.50 | .15 | .44 | .12 |
| | Average | −.52 | .14 | .46 | .13 |
| Average | 25 | −.47 | .12 | .42 | .15 |
| | 50 | −.58 | .12 | .53 | .10 |
| | 75 | −.50 | .12 | .44 | .12 |
| | Average | −.52 | .13 | .46 | .13 |

### 3.2. *Acoustic analysis*

For the acoustic analysis, again the first step has to be a confirmation that our measure for discriminating the sibilants spectrally separates the homorganic conditions sufficiently. Table 5 gives for the Acoustic index the means and standard deviations across speakers for each set at the three timepoints. A repeated measures ANOVA with the within-subjects factors Cluster (/ʃ#ʃ/, /s#s/), Set (CAPS/CASH, MISS/FISH, PARIS/PARISH) and Timepoint (25, 50, 75) was conducted. The main effect Cluster was significant ($F(1,9) = 308.24$, $p < .001$). Neither Set ($F(2,18) < 1$) nor Timepoint ($F(2,18) < 1$) reached significance. As for the EPG data, the Cluster effect shows that our measure successfully separates the two homorganic conditions from each other.[6]

The acoustics confirm the pattern we have observed for the articulatory data. /s#ʃ/ sequences show a strong tendency for regressive assimilation at the 25% analysis timepoint, while /ʃ#s/ sequences show some regressive assimilation, but to a much lesser degree. For the histograms in Figure 5, again the data were divided into ten equally spaced bins at each analysis timepoint and bin centres are marked on the x-axis.

We now turn to the other measurement timepoints. If tokens are produced without assimilation, we expect that over the course of time, for /ʃ#s/ clusters, values should shift from closer to −1 at 25% to closer to 1 at 75%. Vice versa, for a /s#ʃ/ cluster, they should move from closer to −1 at 25% to closer to 1 at 75%. At the

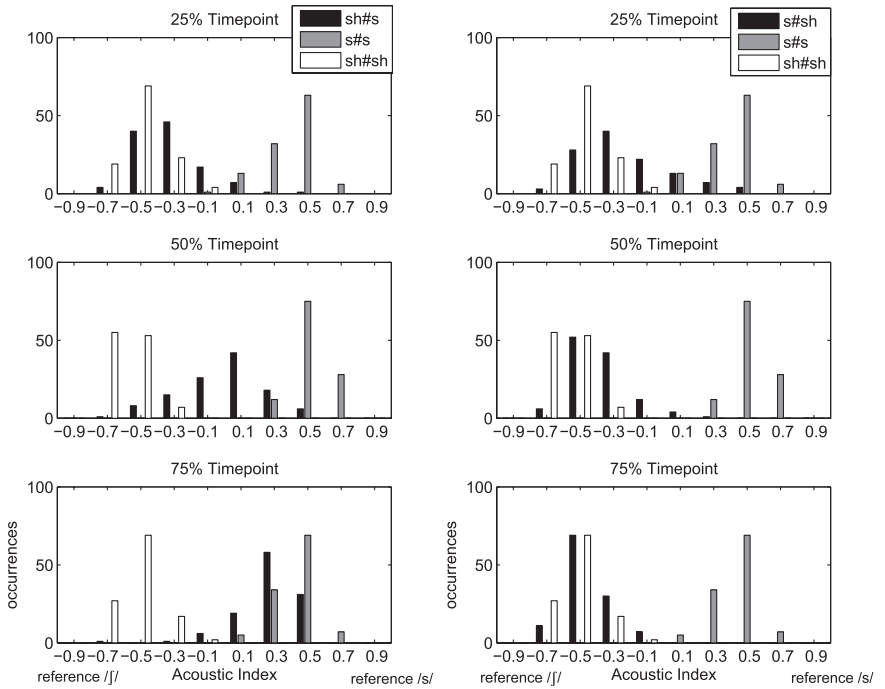


Figure 5. *Comparison of the Acoustic index for the heterorganic clusters and the homorganic control conditions. /ʃ#s/ is compared in the left panels, /s#ʃ/ in the right. The three analysis timepoints are shown top to bottom. Acoustic index values are allocated to ten equally spaced bins. Tick labels on the x-axis indicate bin centres.*

50% timepoint predominantly intermediate values are expected, around zero. Concerning /ʃ#s/, the histograms in the second rows of Figure 3 and Figure 5 evidence a clear change between the 25% and 50% timepoints with tokens migrating towards 1. For /s#ʃ/, on the other hand, not much change is evident. If tokens are already assimilated at 25%, they will not display much change throughout the sibilant sequence.

Note that another difference between the two sequences becomes apparent. For /ʃ#s/, there are some tokens which are still strongly at the /ʃ/ end of the continuum, suggesting the possibility of progressive assimilation. Investigation of the 75% timepoint (third rows of Figure 3, Figure 5) will shed further light on this issue. As expected, /s#ʃ/ sequences do not show much further movement at the 75% timepoint. /ʃ#s/ sequences on the other hand have shown a clear change in distribution from predominantly to the left to predominantly to the right. Note that there are a handful of tokens that display progressive assimilation: At the 75% timepoint, their index values are negative, one token remains within the control distribution of /ʃ/. Note also that more tokens, although positive, are still

close to zero. Comparing this to the /s#ʃ/ sequence, we see a different picture: very few tokens are close to zero, all are below zero and mostly close to the /ʃ/ controls.

### 3.3. *Dynamic analysis*

The analyses so far have revealed that overall, both /s#ʃ/ and /ʃ#s/ assimilate; they differ significantly from both control conditions at the 25% timepoint. Moreover, the /s#ʃ/ condition shows a stronger degree of assimilation because more tokens are in the /ʃ/ control distribution at the 25% timepoint, and the distribution between /s/ and /ʃ/ becomes continuous. For the /ʃ#s/ sequence, some tokens show a tendency towards more positive values, but overall the assimilation is not as strong. The analysis of the timepoints shows that /ʃ#s/ sequences generally show a more dynamic pattern compared to /s#ʃ/ sequences. For the former, many tokens move from a /ʃ/- to a /s/-typical value. For the latter, many tokens are /ʃ/-typical to begin with and become more /ʃ/-typical, if they show any change at all.

This leads to our next point in the analysis, the dynamic aspect of the pattern changes. Do tokens that show in-between values at the 25% timepoint move towards more extreme values or do these articulations retain an 'intermediate' value throughout the sibilant interval? We first discuss the range of change for EPG index values between the 25% and 75% timepoints before showing a series of time-aligned ensemble averages further below. We show in Figure 6 the EPG index

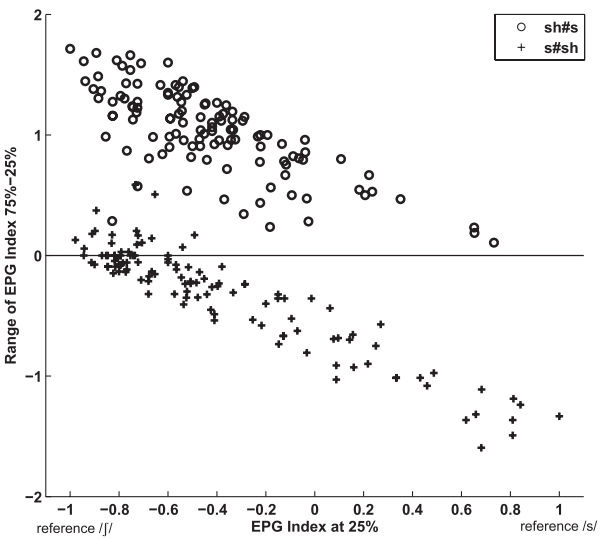


Figure 6. *Scatterplot with data from all subjects showing the range of the EPG index between the 25 and the 75% timepoints plotted against the EPG index at the 25% timepoint.*

value at the 25% timepoint plotted against range of change between the 25% and 75% timepoint (the EPG index at the 25% timepoint was subtracted from the index at the 75% timepoint for each token). This allows us to assess to what degree tokens migrate along the continuum. If articulations do not change, they will be along the horizontal zero line. The maximal value of change is 2 or −2, respectively, since a token can maximally change between the values −1 and 1. Recall that at the 25% timepoint, plotted along the abscissa, we expect unassimilated /ʃ#s/ clusters to approximate −1, while /s#ʃ/ clusters should approximate 1. As already discussed, /s#ʃ/ shows many assimilated tokens at the 25% timepoint, and while /ʃ#s/ sequences also show some assimilation, they do so to a lesser degree. Thus both sibilant sequences have mostly negative values, clustering on the left of the chart. The more strongly assimilated tokens, for both clusters, have a tendency to be close to the horizontal zero line.

Figure 6 shows that for /s#ʃ/ clusters, tokens with an index value of less than 0.5 at the 25% timepoint, show very little or no change at all (those points on or close to the horizontal zero line). Also 'intermediate' productions may not exhibit a large range of change, although there is a clear tendency for less assimilated tokens to show more change. No tokens of the /ʃ#s/ sequence have a range of zero, but three tokens having an index greater than 0.5 approximate the horizontal zero line. These tokens are of particular interest here, since they could be said to display a degree of assimilation comparable to that typically and frequently observed for /s#ʃ/. These three tokens are all *parish soap* tokens from two speakers (SC2, EN3). Auditory analysis and visual inspection of the spectrogram for these three tokens reveal that they were not totally neutralised with *Paris soap*: They have a very brief more /ʃ/-like portion just preceding our 25% analysis timepoint. As an example, Figure 7 displays the spectrogram for speaker EN3 with the 25% analysis timepoint marked by a thick line.

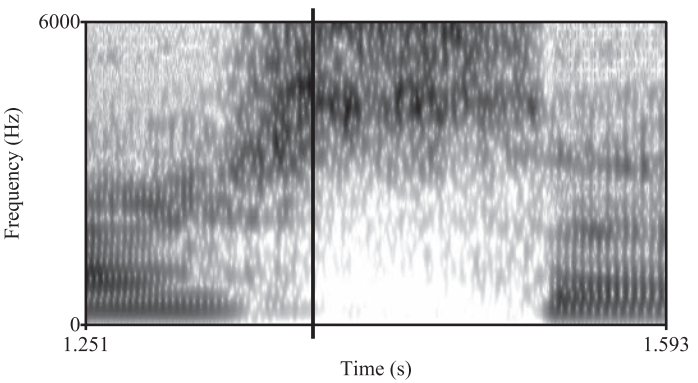


Figure 7. *Spectrogram of EN3 parish soap with 25% analysis timepoint marked by the thick vertical line.*

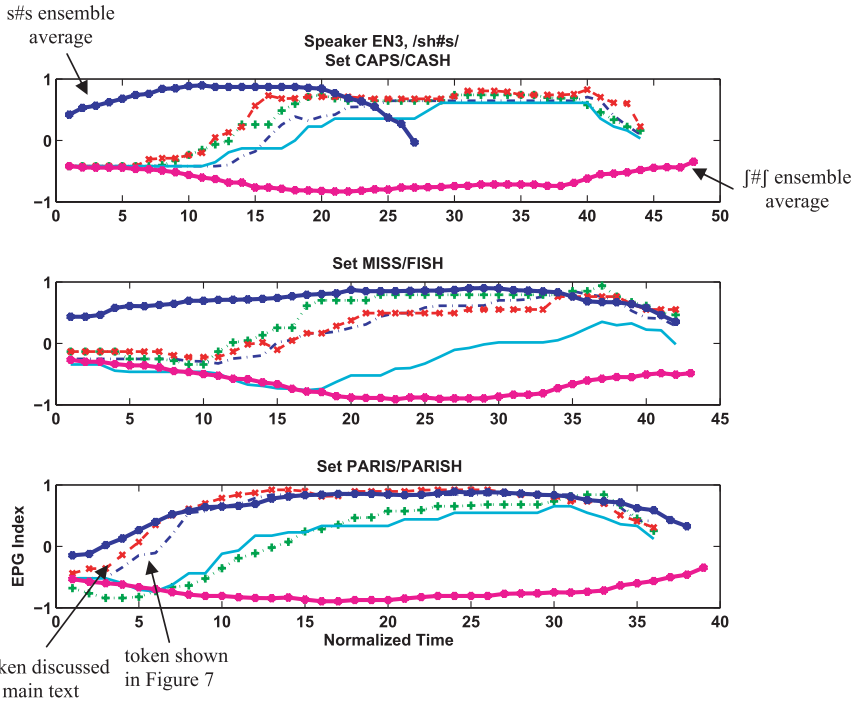


Figure 8. *Time-normalized EPG index for EN3, /ʃ#s/ sequences by set. The parish soap tokens discussed in the main text are in the bottom panel (set PARIS/PARISH) and marked by arrows.*

We moved our analysis timepoint back to 15% for these three tokens and found that the auditorily /ʃʃ/-like portion at the beginning of the fricative sequence is articulatorily intermediate between the two sibilants. The token displayed in Figure 7, for instance, has an EPG index value of −0.14. The other two tokens show values closer to /s/: 0.44 (EN3) and 0.42 (SC2). In Figures 8 and 9 we show time-normalized EPG indices over the course of the sibilant interval for all /ʃ#s/ repetitions for these two speakers, and for reference the ensemble averages for the homorganic conditions. We present the data for all sets to give an impression of the range of variation observed for /ʃ#s/ sequences. Time-normalization was performed separately for each subject and each target sentence. For the four repetitions of a target sentence by a given subject, the average sibilant interval duration was calculated. The repetition that was closest in duration to this average served as the basis to which the other three repetitions were time-aligned. The graphs show the average time-normalized contact pattern for the homorganic conditions with /ʃ#ʃ/ being the bottom dotted line with typical values around −1, and /s#s/ the top dotted line with values around 1. For /ʃ#s/, each individual repetition is shown, separately for each set (since time-normalization was performed separately for each target sentence). The figures show that these assimilated tokens are characterized

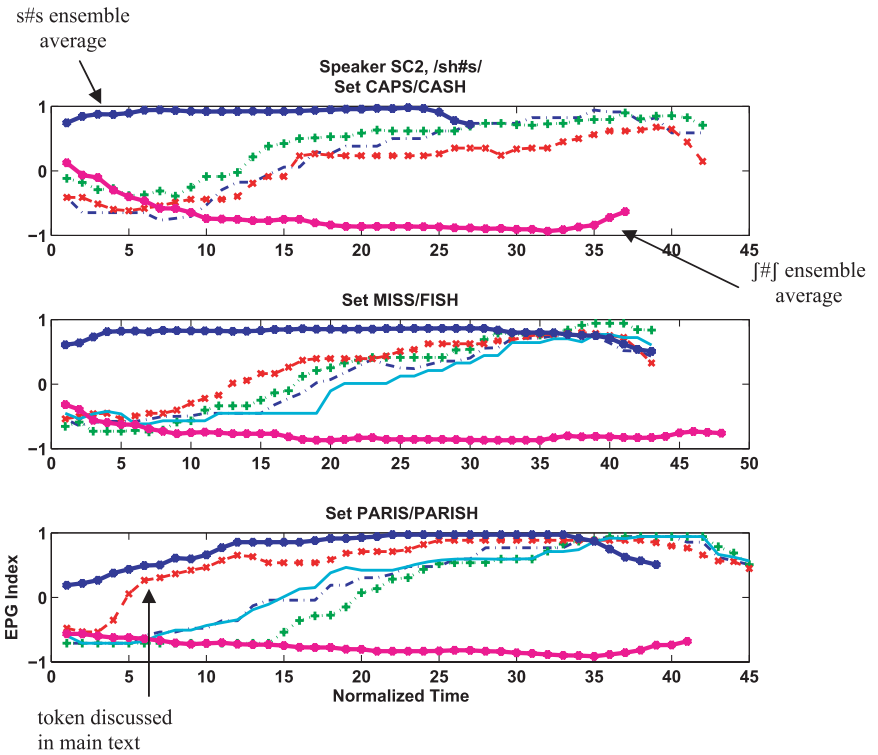


Figure 9. *Time-normalized EPG index for SC2, /ʃ#s/ sequences by set. The parish soap token discussed in the main text is marked in the bottom panel (set PARIS/PARISH).*

by a short /ʃ/-like contact phase at the beginning, but display a very early and rapid change to /s/-typical values: The tokens display strong regressive assimilation.

At this point we also have a closer look at the /ʃ#s/ tokens that seem to display progressive assimilation such that the index value at the 75% point is still negative. We can see in Figures 10 and 11 that this is the case for one *cash Sarah's* token (CAPS/CASH set) and for one *fish soup* token (MISS/FISH set), the latter still remaining within the /ʃ/ control distribution. Now we extend our analysis time-point to later in time, to 85%. The *cash Sarah's* token by SC3 shows an EPG value of −0.15 at 75% and at 85% a value of −0.1 (Figure 10, top panel). The other token, *fish soup* (MISS/FISH set) by EN5, has at 75% a value of −0.54 and at 85% a value of −0.09 (Figure 11, middle panel). The time-normalized plots reveal differences for these two tokens relative to the other repetitions for the respective subject. For speaker SC3 (Figure 10), in the top panel the line marked with an arrow as showing progressive assimilation (the green line in the color version of this Figure) moves gradually toward a more /s/ like value, but fails to ever approach typical /s/ values. Hence this token displays progressive assimilation. Also two of the other

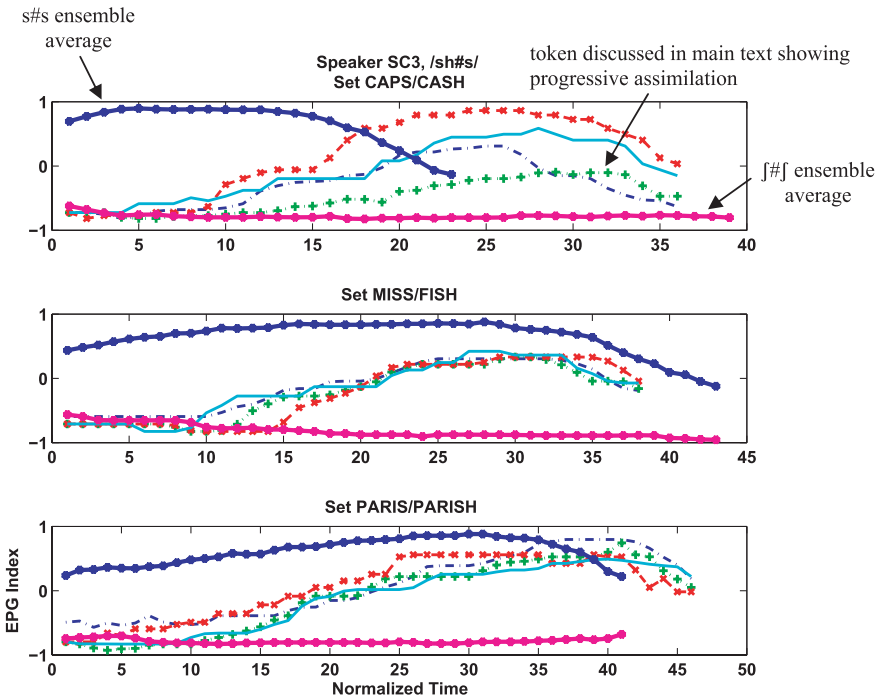


Figure 10. *Time-normalized EPG index for SC3, /ʃ#s/ sequences by set. The token showing progressive assimilation discussed in the text is marked by an arrow.*

repetitions in Figure 10 undershoot the /s/ target. For speaker EN5 (Figure 11, middle panel), a different picture becomes apparent. We see a rapid shift from a /ʃ/- to a /s/-like pattern. The speaker produces a rather long /ʃ/ compared to the other conditions and then changes very rapidly towards /s/, although undershooting the target considerably resulting in a /s/ production intermediate between the /s/ and /ʃ/ reference values.

The time-aligned EPG index curves revealed that /ʃ#s/ tokens with a strong degree of assimilation at the 25% interval still may show a short, relatively less or unassimilated portion of the fricative earlier than the 25% timepoint. A high percentage of /s#ʃ/ sequences was completely assimilated at the 25% timepoint, and hardly displayed any changes in the articulation between the 25% and 75% timepoint analysis (see Figure 6). The question arises whether these tokens may display some dynamic change in the articulation prior to or later than our analysis timepoint. Generally, judging from visual inspection of the time-aligned plots for the /s#ʃ/ sequences for each subject, this is not the case. While /s#ʃ/ sequences may indeed show intermediate EPG index values that change dynamically over the course of the sibilant interval, there are repetitions that are indistinguishable from /ʃ#ʃ/ sequences throughout the sibilant interval. The data of EN4 serve as an ex-

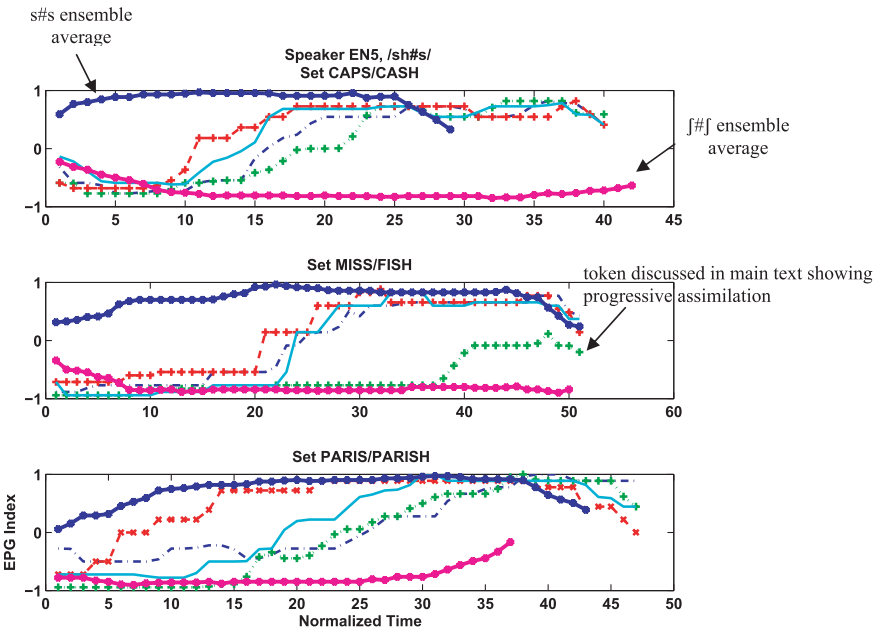


Figure 11. *Time-normalized EPG index for EN5, /ʃ#s/ sequences by set. The token showing progressive assimilation discussed in the text is marked by an arrow.*

ample here (Figure 12). This speaker shows only one intermediate production for the CAPS/CASH set, the other repetitions of that set are unassimilated. For the MISS/FISH and PARIS/PARISH sets, however, all repetitions are completely assimilated throughout the sibilant interval and are, at least by our measures, virtually indistinguishable from the ensemble average for the homorganic /ʃ#ʃ/ condition. For /ʃ#s/, by contrast, we have just seen that there is always a change in the articulation over the course of the sibilant interval, even if only during a very brief period of time or an undershot final target.

Our final analysis concerns the relationship between degree of assimilation and (acoustic) duration of the sibilant interval. As laid out in the Introduction, differences in duration between homorganic (control) sequences and assimilated heterorganic sequences have been used as a pivotal argument in the discussion of the cognitive mechanisms underlying assimilation. A systematic correlation between degree of assimilation and duration is predicted from a gestural perspective, but not necessarily from a feature delinking account of assimilation. Comparisons of surface duration of assimilated forms to homorganic clusters or singleton consonants for any given token are inherently problematic, since many factors influence surface duration simultaneously. We therefore chose to look for a systematic relationship between the duration of the sibilant interval for /s#ʃ/ and /ʃ#s/ sequences and the degree of assimilation at the 25% timepoint. These data are given

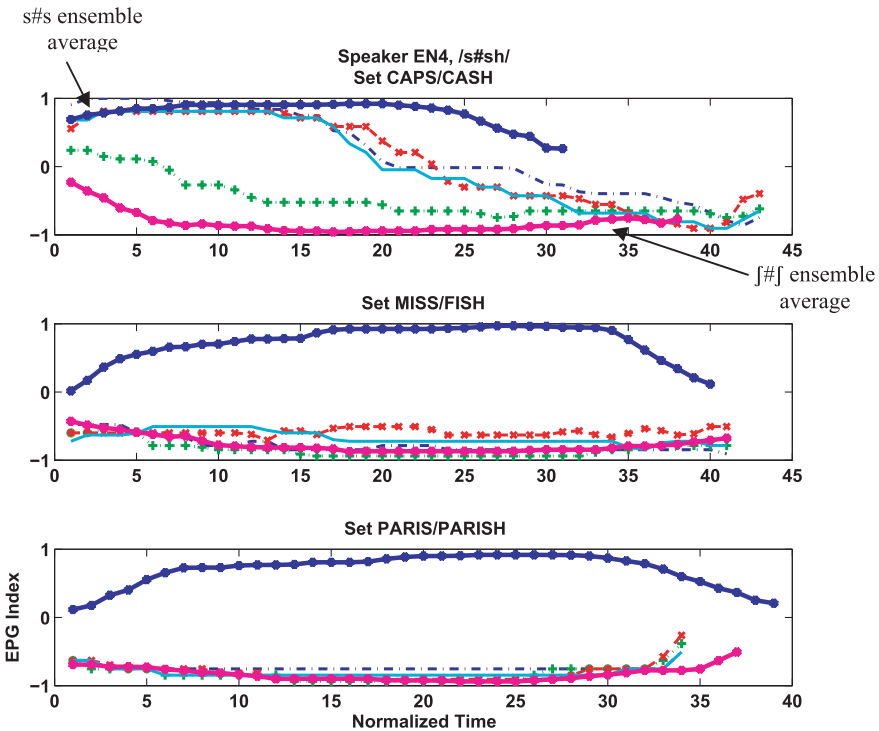


Figure 12. *Time-normalized EPG index for EN4, /s#ʃ/ sequences by set.*

in Figure 13. Duration data were normalized by computing the *z*-score for each subject across all sets and cluster conditions. The CAPS/CASH set is excluded from this analysis, because in this stimulus set, the /s/-final words have a labial preceding the sibilant (*caps*), but the /ʃʃ/-final words do not (*cash*). This means that the acoustic duration of the sibilant interval for conditions with final /s/ (/s#ʃʃ/, /s#s/) is systematically shorter compared to conditions with final /ʃʃ/ (/ʃʃ#s, ʃ#ʃʃ/). We therefore restrict our analysis in this part of the paper to the MISS/FISH and PARIS/PARISH sets. This analysis will not collapse across these two sets, since due to the unstressed syllable in the stimulus words *paris* and *parish,* we observe systematic differences in the duration of the sibilant interval for the two sets.

To establish whether there is a systematic relationship between duration and assimilation, we selected across subjects, but separately for each set, those tokens with duration measures at or below the first quartile (the lower quarter of the data for this measure) and those with duration measures at or above the third quartile (the upper quarter of the data). We then ran a Wilcoxon ranksum test of the EPG indices for those durationally short and long tokens to see whether the EPG index at the 25% timepoint differs significantly for the two duration groups. Vice versa,

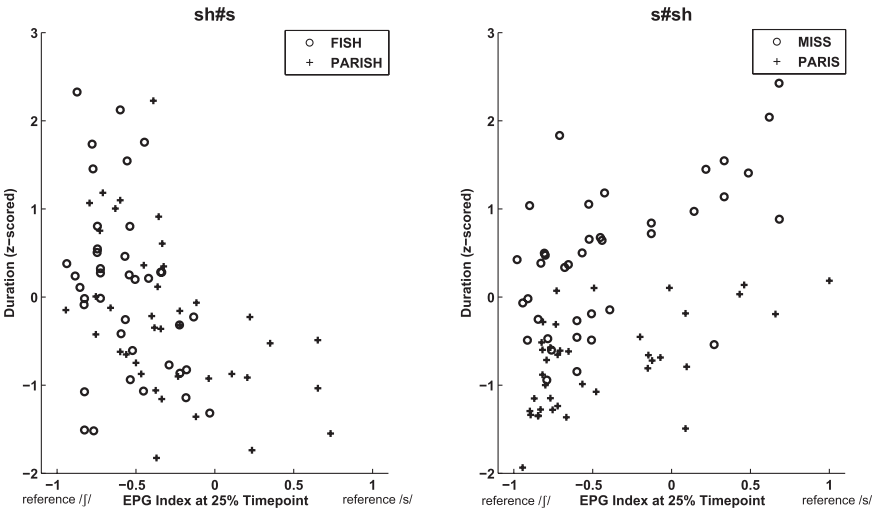


Figure 13. *z-scored duration of sibilant interval plotted against the assimilation index at the 25% timepoint across all subjects for the PARIS|PARISH and MISS|FISH sets. |ʃ#s| is on the left, |s#ʃ| on the right.*

we identified those tokens whose EPG index at the 25% timepoint was below the lower quartile of the overall EPG index distribution, and also those above the upper quartile. Again, a Wilcoxon ranksum test was used to determine whether the 25% most and 25% least assimilated tokens differed significantly in duration (again, this was done across subjects, but separately for each set and sibilant order).

Table 6 gives the mean index and mean duration values for the respective 25% groups. For the /ʃ#s/ order for both sets, the longest tokens have lower EPG index values compared to the shortest tokens (recall that the reference EPG index value for /ʃ/ is −1). Vice versa, for both sets for the /s#ʃ/ order, the longest tokens have

Table 6. *Average EPG index values (25% timepoint) for the tokens with the longest and shortest durations, and average duration values for tokens with the least and most assimilated EPG index values at the 25% timepoint.*

| Sibilant Order | Set | Average EPG Index (25% timepoint) | | Average Duration (*z*-scored) | |
|---|---|---|---|---|---|
| | | longest tokens (highest quantile) | shortest tokens (lowest quantile) | 25% least assimilated tokens | 25% most assimilated tokens |
| ʃ#s | MISS/FISH | −0.68 | −0.43 | 0.35 | −0.47 |
| | PARIS/PARISH | −0.53 | 0.04 | 0.38 | −0.83 |
| s#ʃ | MISS/FISH | 0.11 | −0.59 | 0.59 | 0.18 |
| | PARIS/PARISH | −0.01 | −0.73 | −0.59 | −0.87 |

on average higher index values than the shortest tokens, meaning they are relatively less assimilated (the reference EPG index value for /s/ is 1). This is precisely the pattern to be expected if there is a systematic correlation between degree of assimilation and duration: the more assimilated, the shorter. The duration analysis underscores this result, since all differences are in the expected direction. On average the 25% least assimilated tokens have longer durations compared to the 25% most assimilated tokens. Statistically, all comparisons except for /ʃ#s/ for the MISS/FISH set were significant at the .01 level.

## 4. Discussion

The goal of our study was to investigate the order effect in English sibilant assimilation. Our analysis was able to clearly establish that both sequences, /ʃ#s/ and /s#ʃ/, show assimilation, yet still differ significantly in their assimilatory patterns. A /s#ʃ/ sequence may assimilate rather strongly towards a typical /ʃ/ production, and in some cases be indistinguishable from a /ʃ#ʃ/ production on the measures used. /ʃ#s/ on the other hand shows only a comparatively minor tendency to assimilate towards /s/, and even shows progressive assimilation for some tokens, towards /ʃ/. Taken together, this supports word-final /s/ being more prone to assimilation than word-final /ʃ/ in a sibilant-sibilant context. Assimilated sequences always differ in their EPG and acoustic patterns from /s#s/ sequences. Like many other studies on assimilation, our investigation reveals co-existence of gradient and complete assimilations. Assimilated sequences are significantly shorter compared to non-assimilated sequences, as shown by our duration analysis. This effect was significant except for /ʃ#s/, MISS/FISH set. The /ʃ#s/ MISS/FISH set can be expected to display the least assimilation, since the order /ʃ#s/ is less likely to be assimilated to begin with. There may also be an effect of stress making the unstressed syllable in the words *paris* and *parish* more prone to assimilation compared to the MISS/FISH stimuli (see de Jong, Beckman, and Edwards, 1993, for the effects of stress on coarticulation). Overall a systematic relationship between duration and assimilation could be established. This opens the possibility that complete assimilations are simply the endpoint of a continuum of overlap and reduction. However, this proposal remains unsatisfactory, since the crucial difference between the two sibilant orders that needs to be accounted for lies in the complete assimilation we may observe for /s#ʃ/ (consider Figure 12) but never for the other order. We now consider how perceptual and articulatory factors may condition the asymmetry.

As laid out in the Introduction, asymmetries in assimilation patterns have previously been analysed from a perceptual perspective by positing consonant and context-specific perceptibility scales. However, little can be found about sibilants in this literature, especially in CC clusters. Hura et al. (1992) investigate differences in misperceptions between stops, nasals, and fricatives. They find that of

these three consonant classes, fricatives are most reliably identified correctly, yet their study does not directly target any potential differences among the sibilants. Whalen (1991) finds that in a fricative interval consisting of artificially concatenated /s/ and /ʃ/ portions in either order, the fricatives have the same cross-over points in their perceptual identification curves, confirming that there is no perceptual asymmetry among the two sibilants. Kohler (1990) argues that assimilation occurs for perceptually weak consonants and positions and cites as evidence for this hypothesis that due to their high perceptual salience (German) fricatives do not generally assimilate (see also Jun 2004). In a later textbook publication he notes in passing that specifically /s#ʃ/ sequences in German may in fact assimilate in order to reduce the high degree of articulatory complexity (Kohler 1995). Overall, none of these perception-based accounts can be extended in a non-circular fashion to the asymmetric sibilant assimilation patterns as revealed in the current study.

Resistance to coarticulation has approached asymmetries in assimilatory behaviour from an articulatory perspective. Consonants have been shown to differ in resistance to coarticulation primarily as a function of the control they exert on the tongue dorsum, as well as factors such as manner and mutual compatibility of the demands on the articulators (Recasens et al. 1997; Recasens and Espinosa 2009). Recasens (1999) synthesises a variety of studies on coarticulation, citing several studies in support of the generalization that the extent of carryover versus anticipatory coarticulation varies as a function of tongue dorsum involvement. For consonant vowel coarticulation in VCV sequences, highly constrained consonants show more carryover than anticipatory coarticulation, while less constrained consonants show a predominance of anticipatory coarticulation. The less the tongue dorsum is constrained for a given consonant, the more anticipatory coarticulation will be observed. The higher degree of carryover assimilation is hypothesized to result from the relatively strong control of the dorsum in conjunction with an assumed high degree of biomechanical inertia of the tongue dorsum at the release (cf. also Recasens et al. 1997). To some extent the results of our study are consistent with these studies: /ʃ#s/ showed some progressive assimilation and the articulation during the sibilant interval always changed dynamically from a /ʃ/-like or intermediate value to a /s/-like or intermediate value. That is, there was always evidence for two successive targets, whereas for some tokens of /s#ʃ/ sequences, no such evidence was present and assimilation was strictly regressive. Yet both sequences showed a clear preference for regressive assimilation: There was no clear asymmetry in the direction of assimilation. Rather, the asymmetry lies in the higher propensity for /s/ to assimilate in word-final position and the stochastic occurrence of complete assimilation for only one of the orders (/s#ʃ/). While coarticulation resistance may not suffice to account for the order effect in sibilant assimilation, it may still be the case that differences in articulatory control structures are the basis for the observed pattern. We will now consider in more detail articulatory differences between the sibilants.

The phonological specification of the sibilants as [+/− anterior] implies a relatively similar tongue shape that differs mainly in degree of anteriority. Phonetically, it seems rather to be the case that English /ʃ/ and /s/ are articulated with more global differences in their constrictions, and we believe that these differences may be implicated in the observed asymmetries in behaviour.[7] Despite a similar horizontal location of the narrowest constriction point for the sibilants, the tongue shape behind the constriction differs dramatically (Stone, Faber, Raphael, and Shawker 1992; Toda and Honda 2003; Toda 2009). Several studies show very distinct posterior tongue shapes immediately behind the narrowest constriction point, and suggest an active positioning of the dorsum for /ʃ/, but not for /s/. The dorsal control difference arises from the need to achieve an appropriate cross-sectional tongue shape, constriction length, and distance of the tongue tip to the lower incisors and hence size of the sublingual cavity (Perkell et al. 1979; Stone et al. 1992; Perkell et al. 2004; Toda and Maeda 2006). Data from several further studies are consistent with the view that /ʃ/ exerts a greater dorsal control compared to /s/. Hoole et al. (1993) find for English that the tongue dorsum has more freedom to vary in /s/ compared to /ʃ/, and that /ʃ/ has a stronger tendency to favour carryover (as opposed to anticipatory) coarticulation compared to /s/. Pouplier (2003, 2008) argues in the context of speech errors that sibilants are articulated with different gestures: She suggests that /ʃ/, but not /s/, has a tongue body gesture, while both sibilants control the position of the tongue tip. Further, Stone et al. (1992) investigated cross-sectional tongue shape during the articulation of the different sibilants. They found that the grooving for /s/ was more affected by vowel context compared to /ʃ/. An ultrasound study by Zharkova et al. (2009) suggests that global tongue shape in /s/ is more variable in different vowel contexts compared to /ʃ/, even though a *global* difference in variability was not supported in Hoole et al.'s (1993) EMA study. Further support for the idea that there may be an asymmetry in the control structures underlying the sibilants comes from Perkell et al. (1979). The authors point out that the articulation of /s/ may involve relatively less articulatory precision in horizontal tongue positioning compared to /ʃ/. This is so because /ʃ/ requires a precise control of the distance of the tongue tip to the lower incisors so as to ensure a sufficient size of the sublingual cavity. /s/ on the other hand is articulated with the tongue making contact with the lower incisors and thus allows for a rather "crude" (p. 112) fronting movement, which can probably be taken to mean a ballistic fronting of the tongue.[8] A dorsal constriction is one means by which the distance to the lower incisors can be controlled (see for instance Figure 2e and f in Toda and Honda 2003). Even for subjects who employ a more retroflex rather than a bunched configuration for /ʃ/ (Toda [2009] speaks in the context of French of two different strategies for articulating /ʃ/, and also for English more retroflex articulations can be observed), the tongue still needs to be positioned such that there is a sufficiently long constriction and an appropriate distance to the lower teeth. This would nonetheless imply a tighter tongue shape control for /ʃ/ compared to /s/.

For the current study, all these observations about sibilant articulation may be drawn together and be interpreted to the effect that the different constrictions with which the two sibilants are produced may lead to different consequences of overlap depending on the order. Temporal overlap is intrinsically biased toward C2, since the second consonant will usually have a period where it dominates the vocal tract, after C1 no longer or only weakly exerts an influence on the vocal tract (Ohala 1990). If /ʃ/ and /s/ differ in their control of the tongue body (because /ʃ/ generally imposes more constraints on tongue shape and distance to the lower incisors), a temporal sliding of a tongue tip gesture for /s/ into the more tightly controlled gestural constellation for /ʃ/ (in a /ʃ#s/ sequence) can be assumed to be of comparatively little consequence for the /ʃ/, especially if /ʃ/ controls both the tongue tip and the tongue body, as some studies suggest.[9] If on the other hand, in a /s#ʃ/ sequence, the more strongly controlled tongue body part of /ʃ/ comes to overlap with the less controlled /s/, the area behind the constriction and the distance to the lower incisors would in such a scenario be fully dominated by /ʃ/, leading to strong (near complete) regressive assimilation effects. (Note that even though tongue body differences between the sibilants are posterior to the source, they can nonetheless affect the acoustic output [Perkell et al. 1979: 113]; likewise the difference in constriction length which may be concomittant with the positioning of the tongue body towards the hard palate seems to be an important determinant of the acoustics [Toda and Maeda, 2006]). Small variations in the timing of the overlapping sequence could therefore have relatively stronger consequences for a /s#ʃ/ sequence compared to a /ʃ#s/ sequence. Note though that this account predicts that sequences such as *fish soup* should, in cases of a sufficiently high degree of overlap, be audibly pronounced as similar to *fi[ʃʃ]oup*, but this is certainly an exception, while complete assimilations for /s#ʃ/ are the rule rather than the exception. Rather, it seems that differences in tongue body/shape control may lead to a somewhat greater degree of assimilation for /s#ʃ/ sequences, but other factors are at play in conditioning the occurrence of complete assimilation for that order only.

Our study adds to the patterns observed in many newer studies of assimilation. Both complete and gradient assimilation co-occur regularly in a fashion not predicted by any theoretical account of assimilation in a straightforward way. This leads one to recognize that setting up a dichotomy between cognitive/phonological and physical/phonetic approaches may ultimately not be very useful for understanding assimilatory patterns. A dichotomy is not compatible with fuzzy or nondeterministic interpretations of gradual change from phonetic to phonological distributions. Instead modular theories by their very nature require there to be discrete discontinuities, an assumption for which there is little empirical support (Scobbie 2007). Another approach to the emergence and coexistence of categorical and gradient effects is provided by models of how frequency and probability in language use can influence lexical representations over the course of time. We therefore wish to suggest that while the differences in the articulation of /s/ and /ʃ/ are likely

at the origin of a slight order effect in sibilant assimilation, such starkly different assimilation patterns may be conditioned by the interaction of these articulatory factors with frequency and phonotactic probability effects in language use.

The impact of language use on lexical representations has been captured in exemplar models such as proposed by Pierrehumbert (2001) and Bybee (2000, 2001) as well as in a dynamic field model of lexical representations as recently sketched out by Gafos and Kirov (2009). Based on lenition and related phenomena, this work has shown how the interaction of perception and production through language use can lead to the lexical entrenchment of gradual changes in articulation that are due to variability in how a given token is produced. More frequent and more predictable words are more prone to reduction and lenition compared to low-frequency words. Codas, in particular in word-final position, are generally seen as positions in which consonants (notably stops) undershoot their targets and undergo lenition. Word-final sibilants in English are, however, not usually described as undershooting their targets with respect to place or manner. Rather, we suspect that consonant timing can be affected by frequency, and that this may lead to entrenchment phenomena in consonant timing similar to lenition. If /s#ʃ/ has articulatorily a bias toward stronger assimilation and this order of sibilants is phonotactically more frequent, the interaction of a production bias and phonotactic probability in language use may lead, we argue, to the emergence of an optional allophonic variation (a bimodal probability distribution) for /-s/ final words when followed by /ʃ/. This would obviate the necessity to assume two separate assimilation mechanisms that may lead to gradient or categorical effects respectively. Gradient assimilation may over the course of time, through the shaping force of linguistic experience on lexical representations or their activations, become categorical in that an underlying /ʃ/ is produced instead of a /s/, and the realizations as /ʃ#ʃ/ may appear in a stochastic fashion.

Several studies have shown that collocational frequencies of lexical items determine the phonetic variation with which the words are produced, and have argued that multiword chunks serve as processing units (Gregory, Raymond, Bell, Fosler-Lussier, and Jurafsyk 1999; Bybee 2002; Jaeger and Hoole in press). We suspect that beyond collocational frequency, phonotactic probability is likely to condition similar effects. While there is evidence that certain highly frequent multiword phrases form lexical processing units and may have their own production dynamic, our current data point to a more general effect. It can hardly be argued that the stimuli employed in the current experiment are highly frequent processing chunks. Rather, our data suggest that frequency or phonotactic probability may affect the timing (overlap, coarticulation) of gestures with more frequent sequences being produced with more overlap. While we cannot provide data on phonotactic probability for the two sibilant orders, final /s/ can probably be assumed to be more frequent than /ʃ/ in English due to the plural and the third person singular morpheme endings.[10] It has so far not explicitly been modelled how gestural timing or coupling relations generally may be affected by frequency dynamics. Moreover,

relatively little modelling work has been done on timing relationships between words (but see Smith 1995; Saltzman, Nam, Krivokapic, and Goldstein 2008), but it seems that this is the kind of work needed in order to enhance our models of assimilation. We would like to argue then that we do not need two separate assimilation mechanisms to account for sibilant assimilation, but do need a better understanding of how frequency and probability affect articulatory timing relations within and across words and how language use may feedback to influence intergestural timing, potentially leading to entrenchment and the evolution of allophonic variants.

## 5. Concluding remarks

The fundamental result of our study is that there is a qualitative difference in sibilant assimilation due to the order the sibilants appear in, which cannot be accounted for by any existing single model of word-boundary assimilation in a straightforward fashion. We find that both sequences show regressive assimilation, but /s#ʃ/ sequences may be produced such that they approach forms that are, by our reduced measures, indistinguishable from underlying /ʃ#ʃ/ productions. /ʃ#s/ sequences, on the other hand, always show some evidence of two articulatory targets overlapping/blending to varying degrees. Our proposal is that differences in tongue body control cause a bias for /s#ʃ/ assimilation to be stronger than /ʃ#s/ assimilation. In interaction with a potentially asymmetric phonotactic probability, this may have led to the evolution of an optional occurrence of complete assimilation for final /s/ followed by initial /ʃ/. The data are consistent with the idea of a dynamic relationship between speech production, perception, and cognitive representations, as it is modelled in recent usage-based accounts of consonant lenition. We have argued that these models need to be extended to take into account intergestural timing relationships as part of this dynamic interaction.

## Acknowledgments

We are indebted to our colleagues at QMU who volunteered as subjects and to many of our colleagues in Munich and Edinburgh for discussions and feedback. Work supported by the Deutsche Forschungsgemeinschaft (PO 1269/1-1).

Correspondence e-mail address: pouplier@phonetik.uni-muenchen.de

## Notes

1. We prefer the term ‘complete’ to ‘categorical’, since any absence of differences is always relative to the measures and statistical evaluations employed. By hypothesis, any given measure may allow one to infer that a discrete category change has occurred.

2. Note though that, strictly speaking, symbolic timing slots do not carry any inherent predictions about surface durations, since physical time except for left-to-right sequencing is explicitly not part of this type of representation.
3. A nonlinear phonological account that specifically predicts asymmetries in sibilant assimilation is provided by underspecification theory in which it is assumed that [+anterior] is underspecified (Lahiri 1991). This correctly predicts that assimilation patterns should vary as a function of the order of sibilants, assuming a stipulation that assimilation is exclusively regressive (otherwise both sibilant orders should result in /ʃ#ʃ/, since the specified feature would spread to occupy the underspecified slot in both cases). However, underspecification theory suffers from the same drawbacks as a purely Autosegmental account in assuming assimilation to be a categorical phenomenon by definition.
4. This stimulus set was included to have a set similar to the one employed by Nolan et al. (1996); they used "*claps Shaun, claps Paul, clap Shaun.*"
5. Center of gravity is able to provide a broad classification of the sibilants (Forrest, Weismer, Milenkovic, and Dougall 1988), but has been shown to be less suitable for more fine-grained effects in sibilant production (Jones and Munhall 2003).
6. It will be observed that values of the acoustic index are usually not as close to the extreme values of +/−1 as are the values of the EPG index. This is related to the fact that EPG contact values have only two possible values. Thus it is not unusual for a measured EPG frame to be virtually identical to one of the reference patterns from the homorganic contexts, whereas with the continuous range of values possible for spectral amplitude it will probably be the case that even, for example, a clear /ʃ/ token from a non-homorganic context will differ slightly at every frequency from the amplitude values of the reference /ʃ/ averaged over all homorganic tokens.
7. Rather large inter-speaker differences can be found in the articulation of sibilants; see for instance Toda (2009) or Fletcher and Newman (1991). Whether these differences are causally related to speaker differences in assimilation patterns is unknown.
8. Also for Catalan, Recasens et al. (1997) assume that /s/ in contrast to /ʃ/ does not directly control the position of the tongue body, and it is mainly because of manner that /s/ and /ʃ/ are equal in coarticulation resistance. This suggests similarities to the English sibilant articulation, yet sibilant articulation may differ substantially across languages (Ladefoged and Maddieson 1996; Toda 2009). How this interacts with assimilation patterns is a highly interesting point for future research.
9. For conflicting demands on the same articulator, gestural theory predicts a blended output (Browman and Goldstein 1990b); what exactly that would mean for the tongue tip during overlapping sibilant production is unclear and could probably only be solved by a detailed modelling study. See also Recasens et al. (1997) for an investigation of the effects of articulator compatibility and conflict on coarticulation.
10. A CELEX (Baayen, Piepenbrock, and Gulikers 1995) English word form count reveals that there are over 15 times more /-s/ final compared to /-ʃ/ final word forms, with /-s/ final word forms also having an on average higher token frequency. Comparing the CELEX count for /ʃ-/ and /s-/ initial word forms, we find that there are again about 8 times more /s-/ initial word forms, however, /ʃ-/ initial word forms have on average a higher token frequency than /s-/ initial word forms.

## References

Baayen, R. Harald, Richard Piepenbrock, & L. Gulikers. 1995. *The CELEX lexical database* [CD-Rom]. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.

Barry, Martin C. 1991. Temporal modelling of gestures in articulatory assimilation. *Proceedings of the XIIth International Congress of the Phonetic Sciences, Aix-en-Provence, 4*, 14–17.

Browman, Catherine. 1995. Assimilation as gestural overlap: Comments on Holst and Nolan. In Bruce Connell & Amalia Arvaniti (eds.), *Phonology and Phonetic Evidence. Papers in Laboratory Phonology IV*, 334–342. Cambridge: Cambridge University Press.

Browman, Catherine, & Louis Goldstein. 1990a. Representation and reality: Physical systems and phonological structure. *Journal of Phonetics* 18. 411–424.

Browman, Catherine, & Louis Goldstein. 1990b. Tiers in articulatory phonology, with some implications for casual speech. In John Kingston & Mary E. Beckman (eds.), *Papers in Laboratory Phonology I. Between the Grammar and Physics of Speech*, 340–376. Cambridge: Cambridge University Press.

Browman, Catherine, & Louis Goldstein. 1992. Articulatory phonology: An overview. *Phonetica* 49. 155–180.

Bybee, Joan. 2000. Lexicalization of sound change and alternating environments. In Michael B. Broe & Janet Pierrehumbert (eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, 250–268. Cambridge: Cambridge University Press.

Bybee, Joan. 2001. *Phonology and Language Use*. Cambridge: Cambridge University Press.

Bybee, Joan. 2002. Phonological evidence for exemplar storage of multiword sequences. *Studies in Second Language Acquisition* 24(2). 215–221.

Chitoran, Ioana, & Khalil Iskarous. 2008. Acoustic evidence for high vowel devoicing in Lezgi. *Proceedings of the 8th International Seminar on Speech Production, Strasbourg*, 93–96.

Clements, George N. 1985. The geometry of phonological features. *Phonology Yearbook* 2. 225–252.

de Jong, Kenneth, Mary Beckman, & Jan Edwards. 1993. The interplay between prosodic structure and coarticulation. *Language and Speech* 36(2–3). 197–212.

Ellis, Lucy, & William Hardcastle. 2002. Categorical and gradient properties of assimilation in alveolar to velar sequences: Evidence from EPG and EMA data. *Journal of Phonetics* 30. 373–396.

Fletcher, Samuel G., & Dennis G. Newman. 1991. [s] and [ʃ] as a function of linguapalatal contact place and sibilant groove width. *Journal of the Acoustical Society of America* 89(2). 850–858.

Forrest, Karen, Gary Weismer, Paul Milenkovic, & Ronald Dougall. 1988. Statistical analysis of word-initial voiceless obstruents: Preliminary data. *Journal of the Acoustical Society of America* 84(1). 115–123.

Gafos, Adamantios, & Christo Kirov. 2009. A dynamical model of change in phonological representations: The case of lenition. In François Pellegrino, Egidio Marsico, Ioana Chitoran, & Christophe Coupé (eds.), *Phonological Systems and Complex Adaptive Systems: Phonology and Complexity*, 219–240. Berlin: Mouton de Gruyter.

Goldsmith, John. 1976. Autosegmental phonology. Ph.D. dissertation, MIT.

Goldstein, Louis, Dani Byrd, & Elliot Saltzman. 2006. The role of vocal tract gestural action units in understanding the evolution of phonology. In Michael Arbib (ed.), *From action to language: The mirror neuron system*, 215–249. Cambridge: Cambridge University Press.

Gregory, Michelle L., William D. Raymond, Alan Bell, Eric Fosler-Lussier, & Daniel Jurafsyk. 1999. The effects of collocational strength and contextual predictability in lexical production. *Chicago Linguistics Society* 99. 151–166.

Gusik, Karita M., & Jonathan Harrington. 2007. The quantification of place of articulation assimilation in electropalatographic data using the similarity index (SI). *Advances in Speech-Language Pathology* 9(1). 109–119.

Honorof, Douglas. 1999. Articulatory gestures and Spanish nasal assimilation. Ph.D. dissertation, Yale University.

Hoole, Philip, Noel Nguyen-Trong, & William Hardcastle. 1993. A comparative investigation of coarticulation in fricatives: Electropalatographic, electromagnetic, and acoustic data. *Language and Speech* 36(2–3). 235–260.

Hura, Susan L., Björn Lindblom, & Randy L. Diehl. 1992. On the role of perception in shaping phonological assimilation rules. *Language and Speech* 35(1–2). 59–72.

Jaeger, Marion, & Philip Hoole. In press. Articulatory factors influencing regressive place assimilation across word-boundaries in German. *Journal of Phonetics*.

Jones, Jeffrey A., & Kevin Munhall. 2003. Learning to produce speech with an altered vocal tract: The role of auditory feedback. *Journal of the Acoustical Society of America* 113(1). 532–543.

Jun, Jongho. 1995. Perceptual and articulatory factors in place assimilation: An optimality theoretic approach. Ph.D. dissertation, UCLA, Los Angeles.

Jun, Jongho. 1996. Place assimilation is not the result of gestural overlap: Evidence from Korean and English. *Phonology* 13(3). 377–407.

Jun, Jongho. 2004. Place assimilation. In Bruce Hayes, Robert Kirchner, & Donca Steriade (eds.), *Phonetically Based Phonology*, 58–86. Cambridge: Cambridge University Press.

Kochetov, Alexei, & Marianne Pouplier. 2008. Phonetic variability and grammatical knowledge. An articulatory study of Korean place assimilation. *Phonology* 25. 399–431.

Kohler, Klaus. 1990. Segmental reduction in connected speech in German: Phonological facts and phonetic explanations. In William J. Hardcastle and Alain Marchal (eds.), *Speech Production and Speech Modelling*, 69–92. Dordrecht: Kluwer.

Kohler, Klaus. 1995. *Einführung in die Phonetik des Deutschen*. Berlin: Erich Schmidt.

Kühnert, Barbara, & Philip Hoole. 2004. Speaker-specific kinematic properties of alveolar reductions in English and German. *Clinical Linguistics and Phonetics* 18(6–8). 559–575.

Ladefoged, Peter, & Ian Maddieson. 1996. *The Sounds of the World's Languages*. Oxford: Blackwell.

Lahiri, Aditi. 1991. Anteriority in sibilants. *Proceedings of the XII International Congress of the Phonetic Sciences, Aix-en-Provence* 1. 348–388.

McCarthy, John J. 1988. Feature geometry and dependency: A review. *Phonetica* 45. 84–108.

Niebuhr, Oliver, Leonardo Lancia, & Christine Meunier. 2008. On place assimilation in French sibilant sequences. *Proceedings of the 8th International Seminar on Speech Production, Strasbourg*.

Nolan, Francis. 1992. The descriptive role of segments: Evidence from assimilation. In Gerard J. Docherty & D. Robert Ladd (eds.), *Laboratory Phonology II: Gesture, Segment, Prosody*, 261–280. Cambridge: Cambridge University Press.

Nolan, Francis. 1999. The devil is in the detail. *Proceedings of the XIVth International Congress of the Phonetic Sciences, San Francisco*, 1–8.

Nolan, Francis, Tara Holst, & Barbara Kühnert. 1996. Modelling [s] to [ʃ] accomodation in English. *Journal of Phonetics* 24. 113–137.

Ohala, John. 1990. The phonetics and phonology aspects of assimilation. In John Kingston and Mary E. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, 258–275. Cambridge: Cambridge University Press.

Perkell, Joseph S., Suzanne E. Boyce, & Kenneth Stevens. 1979. Articulatory and acoustic correlations of the [s-ʃ] distinction. *Speech Communication Papers Presented at the 97th Meeting of the Acoustical Society of America*, 109–113.

Perkell, Joseph S., Melanie Matthies, Mark Tiede, Harlan Lane, Majid Zandipour, & Nicole Marrone. 2004. The distinctness of speakers' /s/-/ʃ/ contrast is related to their auditory discrimination and use of an articulatory saturation effect. *Journal of Speech, Language, and Hearing Research* 47(6). 1259–1269.

Pierrehumbert, Janet. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In Joan H. Bybee & Paul Hopper (eds.), *Frequency and the Emergence of Linguistic Structure*, 137–157. Amsterdam: John Benjamins.

Pouplier, Marianne. 2003. Units of phonological encoding: Empirical evidence. Ph.D. dissertation, Yale University.

Pouplier, Marianne. 2008. The role of a coda consonant as error trigger in repetition tasks. *Journal of Phonetics* 36. 114–140.

Recasens, Daniel. 1999. Lingual coarticulation. In William Hardcastle and Nigel Hewlett (eds.), *Coarticulation: Theory, Data, and Techniques*, 80–104. Cambridge: Cambridge University Press.

Recasens, Daniel. 2006. Integrating coarticulation, blending and assimilation into a model of articulatory constraints. In Louis Goldstein, Douglas Whalen, & Catherine Best (eds.), *Laboratory Phonology 8*, 611–634. Berlin and New York: Mouton de Gruyter.

Recasens, Daniel, & Aina Espinosa. 2009. An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan. *Journal of the Acoustical Society of America* 125(4). 2288–2298.

Recasens, Daniel, Jordi Fontdevila, & Maria Dolors Pallarès. 1992. Alveolar-palatal correlations in coarticulatory activity for a selected group of Catalan consonants. *Bulletin de la Communication Parlée* 2. 59–72.

Recasens, Daniel, Maria Dolors Pallarès, & Jordi Fontdevila. 1997. A model of lingual coarticulation based on articulatory constraints. *Journal of the Acoustical Society of America* 102(1). 544–561.

Saltzman, Elliot, Hosung Nam, Jelena Krivokapic, & Louis Goldstein. 2008. A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. In Plínio A. Barbosa & Sandra Madureira (eds.), *Proceedings of the Speech Prosody 2008 Conference, Campinas, Brazil*, 175–184.

Scobbie, James. 1995. What do we do when phonology is powerful enough to imitate phonetics? Comments on Zsiga. In Bruce Connell & Amalia Arvaniti (eds.), *Papers in Laboratory Phonology IV: Phonology and Phonetic Evidence*, 303–314. Cambridge: Cambridge University Press.

Scobbie, James. 2007. Interface and overlap in phonetics and phonology. In Gilian Ramchand & Charles Reiss (eds.), *The Oxford Handbook of Linguistic Interfaces*, 17–52. Oxford: Oxford University Press.

Scobbie, James, & Marianne Pouplier. 2010. Conditioning factors in external sandhi: An EPG study of vocalisation and retraction of word-final English /l/. *Journal of Phonetics* 38(2). 240–259.

Smith, Caroline. 1995. Prosodic patterns in the coordination of vowel and consonant gestures. In Bruce Connell & Amalia Arvaniti (eds.), *Phonology and Phonetic Evidence. Papers in Laboratory Phonology IV*, 205–222. Cambridge: Cambridge University Press.

Son, Minjung, Alexei Kochetov, & Marianne Pouplier. 2007. The role of gestural overlap in perceptual place assimilation in Korean. In Jennifer Cole & José Ignacio Hualde (eds.), *Papers in Laboratory Phonology IX*, 507–534. Berlin: Mouton de Gruyter.

Steriade, Donca. 2001. Directional asymmetries in place assimilation: A perceptual account. In Elizabeth Hume & Keith Johnson (eds.), *The Role of Speech Perception in Phonology*, 219–250. San Diego: Academic Press.

Steriade, Donca. 2009. The phonology of perceptibility effects: The P-map and its consequences for constraint organization. In Sharon Inkelas & Kristin Hanson (eds.), *On the Nature of the Word. Studies in Honor of Paul Kiparsky*, 151–180. Cambridge, MA: MIT Press.

Stone, Maureen, Alice Faber, Lawrence J. Raphael, & Thomas Shawker. 1992. Cross-sectional tongue shape and linguopalatal contact patterns in [s], [ʃ], and [l]. *Journal of Phonetics* 20. 253–270.

Thomson, D. J. 1982. Spectrum estimation and harmonic analysis. *Proceedings of the IEEE* 70(9). 1055–1096.

Toda, Martine. 2009. Etude articulatoire et acoustique des fricatives sibilantes. Ph.D. thesis, Université Paris 3.

Toda, Martine, & Kiyoshi Honda. 2003. An MRI based cross-linguistic study of sibilant fricatives. *Proceedings of the 6th International Seminar on Speech Production, Sydney*, 290–295.

Toda, Martine, & Shinji Maeda. 2006. Quantal aspects of non anterior sibilant fricatives: A simulation study. *Proceedings of the 7th International Seminar on Speech Production, Ubatuba, Brazil*, 573–580.

Whalen, Douglas H. 1991. Perception of the English /s/-/ʃ/ distinction relies on fricative noises and transitions, not on brief spectral slices. *Journal of the Acoustical Society of America* 90(4). 1776–1785.

Zharkova, Natalia, Nigel Hewlett, & William J. Hardcastle. 2009. An ultrasound study of lingual coarticulation in children and adults. *Full Research Report ESRC End of Award Report, RES-000-22-2833. Swindon: ESRC.*

Zsiga, Elizabeth. 1995. An acoustic and electropalatographic study of lexical and postlexical palatalization in American English. In Bruce Connell & Amalia Arvaniti (eds.), *Phonology and Phonetic Evidence. Papers in Laboratory Phonology IV*, 282–302. Cambridge: Cambridge University Press.